Utilizing large Ribosomal Subunit protein L3 for phylogenetic identification of novel bacterial lineages in Zodletone Spring.

Sam Kimbrough, Ibrahim Farag, Noha Youssef, Mostafa Elshahed

Abstract

Zodletone Spring is a sulfur and sulfide rich anoxic spring in southwestern Oklahoma characterized by a high level of phylogenetic diversity. We utilized a genome-resolved metagenomics approach to recover genomes of uncultured microbial lineages in the spring. Sediments from the source of the spring were collected, and 551 Gbp of metagenomics raw data were obtained using paired-end Illumina sequencing. The raw data was assembled and genomic bins were recovered from the dataset. A total of 380 genomes with various completion levels were recovered. 321 of these genomes were from the kingdom Bacteria and 59 were from the kingdom Archaea. We used ribosomal protein L3 (RPL3) as a phylogenetic marker to identify genomes belonging to novel lineages within this large dataset. My research focused on the identification of novel class and order level lineages within the phylum Chloroflexi and the classes Deltaproteobacteria (Phylum Proteobacteria), and Clostridia (Phylum Firmicutes). Our analysis putatively identified 19 genomes belonging to 9 novel lineages within the Chloroflexi, 12 genomes defining 8 novel lineages within the Deltaproteobacteria, and 3 genomes belonging to 3 novel lineages within the Clostridia. Future plans include detailed metabolic characterization of these lineages to better understand their metabolic abilities, physiological preferences, and ecological role in Zodletone spring.

Introduction

The vast majority of organisms residing in a specific environment are either not cultivatable in a laboratory setting (Rinke et al.), or grow at such a slow pace that they are not a reasonable candidate for growth in a laboratory (Hoehler and Jørgensen). Thankfully, our capability to characterize the phylogenic relationships, metabolic capabilities, and ecological roles and preferences of uncultured organisms in a culture-independent manner using genomic approaches is greatly expanding. Sequencing technologies are rapidly improving and becoming more accessible (Mardis). An entire array of metagenomics assembly tools are freely available for a variety of sample types (Vollmers et al.), as are tools for the binning of metagenomes (Kang et al.; Wu et al.). Additionally, software for the creation of phylogenetic trees from the large datasets metagenomics surveys can create is also improving greatly (Stamatakis).

Zodletone spring is an anoxic, sulfur- and sulfide rich, open air spring located in southwestern Oklahoma. Previous studies have indicated a large and diverse microbial community is present within the spring (Elshahed et al.; Luo et al.). Within the microbial communities at large, the bacterial communities themselves have proven to be quite diverse. The presence of sulfur- and sulfate-reducing lineages in addition to novel candidate divisions (Elshahed et al.).

Bioinformatics and its approaches tend to rely on the small subunit 16S rRNA gene (16S) for identification of genomic bins has been shown to be an unreliable method as it can be difficult to properly align while having a tendency to mischaracterize lineages. Gene-coding sequences, however, are highly conserved amongst organisms and prove to be much easier to properly align. Ribosomal proteins, such as ribosomal protein L3 (RPL3), are an example of such sequences.

Here, my goal was to further characterize the bacterial communities of Zodletone Spring using RPL3-based phylogenetic analysis of genomes assembled using genome-resolved metagenomics. It is our hope to demonstrate rpL3 and other conserved gene-coding sequences can be a useful phylogenetic marker. Furthermore, we hope to uncover additional novel microbial lineages living within the Spring.

Methods

1. Sampling.

Samples were obtained from Zodletone spring in Fall 2015. Sediment samples were scooped from the source of the spring and placed in 50 ml falcon tube. Water samples were collected in 10 liter plastic containers. The samples were placed on ice and transferred to the laboratory (160 miles) were they were promptly stored at -20^{0}C. Water samples were concentrated by centrifugation and resuspension into sterile water. DNA was extracted from sediment samples and concentrated water samples using commercially DNA extraction kits. Qiagen. Sequencing was conducted using the services of a commercial provider (Novogenes, Beijing, China).

2. Metagenomic datasets quality filtration, assembly and binning. Zodletone metagenomic raw, unassembled fastq reads were preprocessed using Trimmomatic tool (Bolger et al.) to remove Illumina adapters. Then, the reads were filtered out based on their qualities using the adaptive trimming tool Sickle (Joshi and Fass) applying the default parameters. High quality fastq reads were de-novo assembled using Megahit v1.1.1 (Li et al.) applying the following parameters: --presets 'meta-large' --kmin-1pass --mink 27, --maxk 97, --step 10, --min_contig 1000. Two types of Zodletone samples were collected and sequenced, water and sediment

samples. This method section discusses our efforts to *in-silico* analyze Zodletone sediment samples.

Assembled Contigs with minimum lengths ≥1000 nts were binned based on the read coverage information and phylogenetic affiliations of the analyzed contigs using the Maxbin v2.2.1 (Wu et al.) program applying the default settings. The quality of the binned draft genomes, completeness and strain heterogeneity were further assessed using CheckM (Parks et al.). Binned genomes showing contamination levels higher than 15% and/or strain heterogeneity more than 10% were further refined based on their sequence composition, tetranucleotide frequencies and GC contents. We attempted to salvage these low quality bins using nonlinear dimension reduction algorithm BH-SNE based software, Vizbin (Laczny et al.). Genome drafts remained with poor quality after the refinement steps were removed.

3. Draft genomes phylogenetic assignment based on a single phylomarker gene approach

A set of 15 different single copy phylogenetic marker genes were selected to conduct the preliminary phylogenetic assignments of the draft genomes recovered from Zodletone Sediment (RPL2, RPL3, RPL4, RPL5 RPL6, RPL14, RPL15, RPL18, RPL22, RPL24, RSSU3, RSSU8, RSSU10, RSSU17 and RSSU19). We observed that the phylomarker RPL3 was well represented in most of the sediment recovered draft genomes (190/321), followed by RPL11 (99/321). We proceeded with these two phylomarkers to comprehensively evaluate the phylogenetic positions of the sediment draft genomes.

In-house RPL3 and RPL11 protein databases were created using HMMer v3.1b2-mpi (Finn et al.), and based on Pfam seed alignments for protein families PF00297 and PF00298, respectively. These databases were used to search the sediment draft genomes for RPL3 and

RPL11 protein sequences using HMMer tool and through applying the default parameters. To confirm that the extracted RPL3 and RPL11 protein sequences are genuine RPL3 and RPL11 sequences, we blasted those sequences against NCBI nr database and all non-RPL3 and RPL11 sequences were removed from any subsequent analyses. The extracted sequences were used for preliminary phylogenetic identification of bins by blasting them against the UniRef100 database (Suzek et al.).

In parallel, two different phylogenetic trees were created for the phylomarkers RPL3 and RPL11 protein sequences. For each tree, the extracted phylomarkers from the Zodletone draft genomes were aligned together with markers recovered from various reference genomes representing 380 bacterial and archaeal phyla. The sequences were aligned using Muscle v3.8.31 (Edgar) applying the default parameters. The quality of the alignment was checked using Jalview v2 (Waterhouse et al.)

Finally, We created the RPL3 and RPL11 based tress following Maximum Likelihood approach using RAxML v8.28 (Stamatakis) and applying the following parameters -m PROTGAMMABLOSUM62 -f a -p 12345 -x 12345 -# 100 -o Pectate_lyase -T 12.

4. Sub-class phylogenetic evaluation of the Deltaproteobacteria-affiliated draft genomes

The initial phylogenetic assignments of the Zodletone draft genomes showed the presence of 45 genomes potentially belonging to class Deltaproteobacteria. In this analysis, we attempted to accurately assign these genomes to their respective Deltaproteobacteria sub-class levels using the phylomarker RPL3. The sub-class phylogenetic assignments was achieved through aligning the RPL3 protein sequences extracted from the draft genomes together with 37 Deltaproteobacteria reference RPL3 protein sequences, representing 10 different orders with

Deltaproteobacteria. The alignment and tree construction were performed as described in section 3. Any trees not aligned with Deltaproteobacteria would be removed from further analysis.

5. Sub-class phylogenetic evaluation of the Chloroflexi-affiliated draft genomes.

Initial phylogenetic assignments of Zodletone draft genomes showed the presence of 25 genomes potentially belonging to phylum Chloroflexi. Analysis was conducted as described in Section 4 with 72 reference RPL3 protein sequences representing 12 families.

6. Sub-class phylogenetic evaluation of the Clostridia-affiliated draft genomes.

Initial phylogenetic assignments of Zodletone draft genomes indicated the presence of 6 genomes potentially belonging to the class Clostridia. Analysis was conducted as described in Section 4 with 132 reference RPL3 protein sequences representing 17 orders.

Results

1. General

380 total genomes were recovered from the Zodletone sediment samples. The samples fell into one of two groups. 321 of the samples were determined to be Bacteria while the remaining 59 were determined to be Archaea (Figure 1).

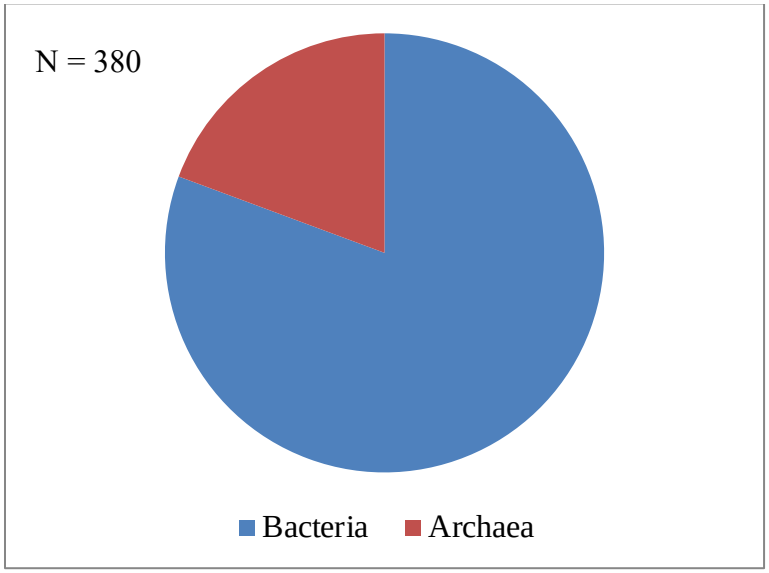

Figure 1. Kindgom level phylogenetic distribution of Zodletone sediment samples

2. Deltaproteobacteria.

Out of the 321 bacterial genomes from the Zodletone sediment samples, 16 were identified using BLAST (Table 1) and phylogenetic tree construction as Deltaproteobacteria. Our focus was on the 12 genomes that were determined to constitute 8 novel clades through phylogenetic affiliation, heretofore referred to as δ1-δ8 (Figure 2). Other genomes were present, but clustered with known Deltaproteobacteria. Half of the samples had no strong phylogenetic affiliation while others could be putatively classified (Figure 3).

Bin Name	Novel Clade Designation	Closest Relative	Closest Relative NCBI Accession	Identity Match	Putative Order
Zodletone_maxbin.out.0082_L3	δ1	*Anaerostipes hadrus*	WP_055257643.1	52%	Unknown
Zodletone_maxbin.out.0096_L3	δ2	*Syntrophorhabdus aromaticivorans*	WP_028894405.1	50%	Unknown
Zodletone_maxbin.out.0241_L3	δ2	*Ruminococcus albus*	WP_024858070.1	53%	Unknown
Zodletone_maxbin.out.0101_L3	δ3	*Haliangium ochraceum*	WP_096058756.1	51%	Myxococcales
Zodletone_maxbin_2.out.041_L3	δ4	*Sandaracinus amylolyticus*	WP_053234415.1	64%	Myxococcales
Zodletone_maxbin.out.0505_L3	δ5	*Pelobacter seleniigenes*	WP_029911168.1	83%	Desulfuromonadales
Zodletone_maxbin.out.0553_L3	δ6	*Desulfospira joergensenii*	WP_033398270.1	56%	Desulfobacterales
Zodletone_maxbin.out.0063_L3	δ7	*Desulfovibrio mexicanus*	WP_089271658.1	77%	Desulfovibrionales
Zodletone maxbin.out.0050_L3	δ8	*Desulfuromonas thiophila*	WP_092076091.1	55%	Unknown
Zodletone maxbin.out.0088_L3	δ8	*Pyramidobacter piscolens*	WP_009165424.1	55%	Unknown
Zodletone maxbin.out.0311_L3	δ8	*Desulfuromonas acetoxidans*	WP_040367924.1	55%	Unknown
Zodletone_maxbin.out.0673_L3	δ8	*Desulfuromonas thiophila*	WP_092076091.1	55%	Unknown
Zodletone_maxbin.out.0466_1_L3	N/A	*Desulfuromonas soudanensis*	WP_053551583.1	51%	Unknown
Zodletone_maxbin.out.0122_L3	N/A	*Desulfuromonas soudanensis*	WP_053551583.1	70%	Desulfuromonadales
Zodletone_maxbin.out.0133_1_L3	N/A	*Desulfatiglans anilini*	WP_028322657.1	68%	Desulfobacterales
Zodletone_maxbin.out.0248_L3	N/A	*Syntrophorhabdus aromaticivorans*	WP_028894405	60%	Syntrophobacterales

Table 1. BLAST results of Deltaproteobacteria-associated Zodletone samples and putative affiliation derived from clustering in phylogenetic tree.

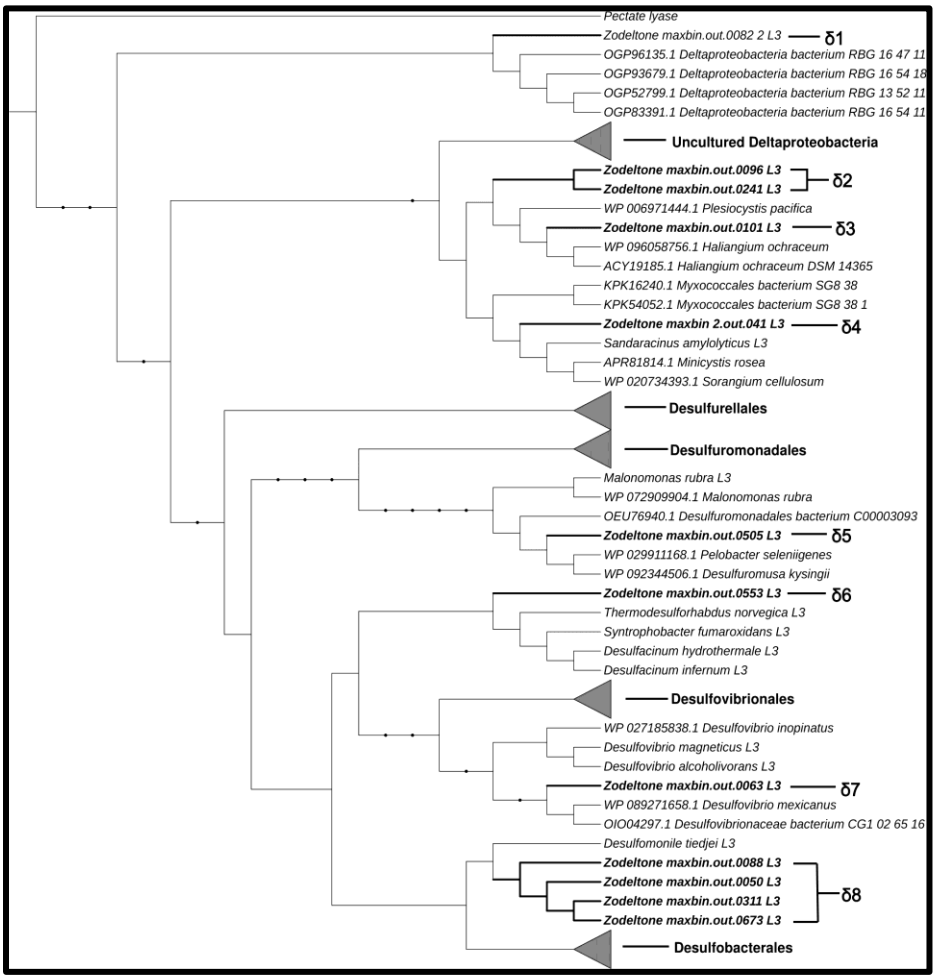

Figure 2. Phylogenetic tree of Deltaproteobacteria-associated Zodletone RPL3 sequences and associated reference RPL3 sequences retrieved from the NCBI database.

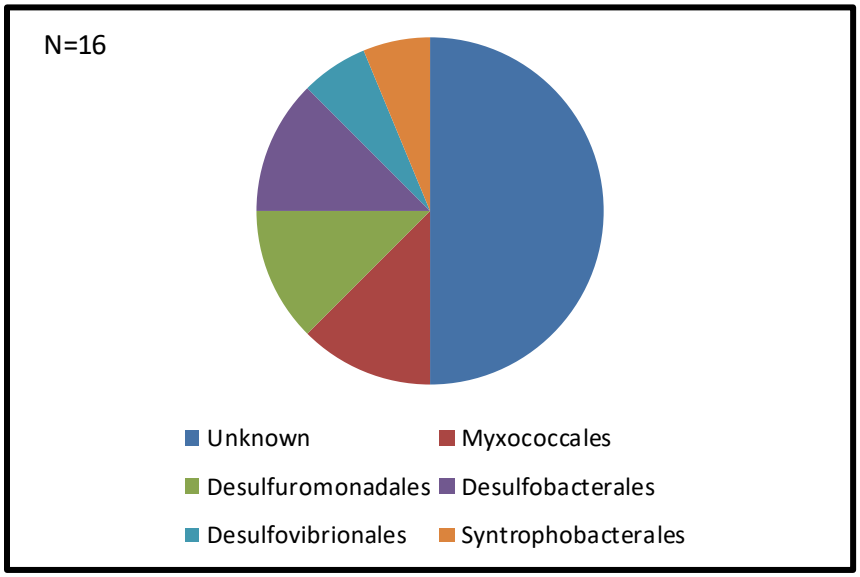

Figure 3. Phylogenetic assignments of Deltaproteobacteria-associated Zodletone samples.

3. Chloroflexi.

Of the 321bacterial genomes obtained from Zodletone sediment samples, 24 were determined to belong to the phylum Chloroflexi using BLAST and phylogenetic tree affiliation (Table 2). 19 genomes were determined to belong to 9 novel clades, heretofore referred to as Chl1-9 (Figure 4). The remaining five genomes were found to cluster with cultured members of Chloroflexi. The phylogenetic distribution of samples can be seen in Figure 5.

Bin Name	Novel Clade Designation	Closest Relative	Closest Relative NCBI Accession	Identity Match	Putative Class
Zodletone_maxbin.out.0001_L3	Chl1	*Anaerolinea thermophila*	WP_013559429.1	78%	Anaerolineae
Zodletone_maxbin.out.0079_1_L3	Chl2	*Anaerolineaceae* bacterium 4572_5.2	OQY37026.1	76%	Anaerolineae
Zodletone_maxbin.out.0356_L3	Chl2	*Chloroflexi* bacterium RBG_13_60_9	OGO07695.1	89%	Anaerolineae
Zodletone_maxbin.out.0208_L3	Chl2	*Chloroflexi* bacterium RBG_13_60_9	OGO07695.1	95%	Anaerolineae
Zodletone_maxbin.out.0230_L3	Chl3	*Chloroflexi* bacterium	PJF22717.1	70%	Caldilineae
Zodletone_maxbin.out.0105_1_L3	Chl3	*Chloroflexi* bacterium	PJF22717.1	66%	Caldilineae
Zodletone_maxbin.out.0183_1_L3	Chl3	*Chloroflexi* bacterium	PJF22717.1	63%	Caldilineae
Zodletone_maxbin.out.0426_1_L3	Chl4	*Candidatus Chloroploca asiatica*	WP_097650759.1	61%	Caldilineae
Zodletone_maxbin.out.0335_L3	Chl4	*Chloroflexi* bacterium	PJF44882.1	59%	Caldilineae
Zodletone_maxbin_2.out.0049_L3	Chl4	*Chloroflexi* bacterium	PJF44882.1	60%	Caldilineae
Zodletone_maxbin.out.0274_L3	Chl5	*Chloroflexi* bacterium	PJF22717.1	62%	Ardenticatenia
Zodletone_maxbin_2.out.018_L3	Chl5	*Chloroflexi* bacterium	PJF22717.1	63%	Ardenticatenia
Zodletone_maxbin.out.0035_L3	Chl6	*Chloroflexi* bacterium RBG_19FT_COMBO_48_23	OGO60502.1	92%	Uncultured
Zodletone_maxbin.out.0226_1_L3	Chl7	*Chloroflexi* bacterium RBG_13_46_14	OGN87136.1	80%	Uncultured
Zodletone_maxbin.out.0226_2_L3	Chl7	*Chloroflexi* bacterium RBG_13_46_14	OGN87136.1	78%	Uncultured
Zodletone_maxbin.out.0383_1_L3	Chl8	*Chloroflexi* bacterium RBG_13_56_8b	OGO06670.1	72%	Uncultured
Zodletone_maxbin.out.0417_L3	Chl8	*Chloroflexi* bacterium RBG_13_46_14	OGN87136.1	72%	Uncultured
Zodletone_maxbin.out.0144_L3	Chl9	*Dehalogenimonas alkenigignens*	WP_058438097.1	67%	Dehalococcoidia
Zodletone_maxbin.out.0330_1_L3	Chl9	*Dehalogenimonas sp. WBC-2*	AKG53164.1	69%	Dehalococcoidia
Zodletone_maxbin.out.0193_L3	N/A	*Anaerolineae* bacterium SM23_ 63	KPK93550.1	78%	Anaerolineae
Zodletone_maxbin.out.0303_L3	N/A	*Anaerolineae* bacterium UTCFX2	OQY87237.1	83%	Anaerolineae
Zodletone_maxbin.out.0280_1_L3	N/A	*Chloroflexi* bacterium RBG_16_58_14	OGO41932.1	94%	Uncultured
Zodletone_maxbin.out.0446_L3	N/A	*Dehalococcoidia* bacterium DG_18	KPJ54579.1	73%	Dehalococcoidia
Zodletone_maxbin.out.0626_1_L3	N/A	*Dehalogenimonas sp. WBC-2*	AKG53164.1	69%	Dehalococcoidia

Table 2. BLAST results of Chloroflexi-associated Zodletone samples and putative affiliation derived from clustering in phylogenetic tree.

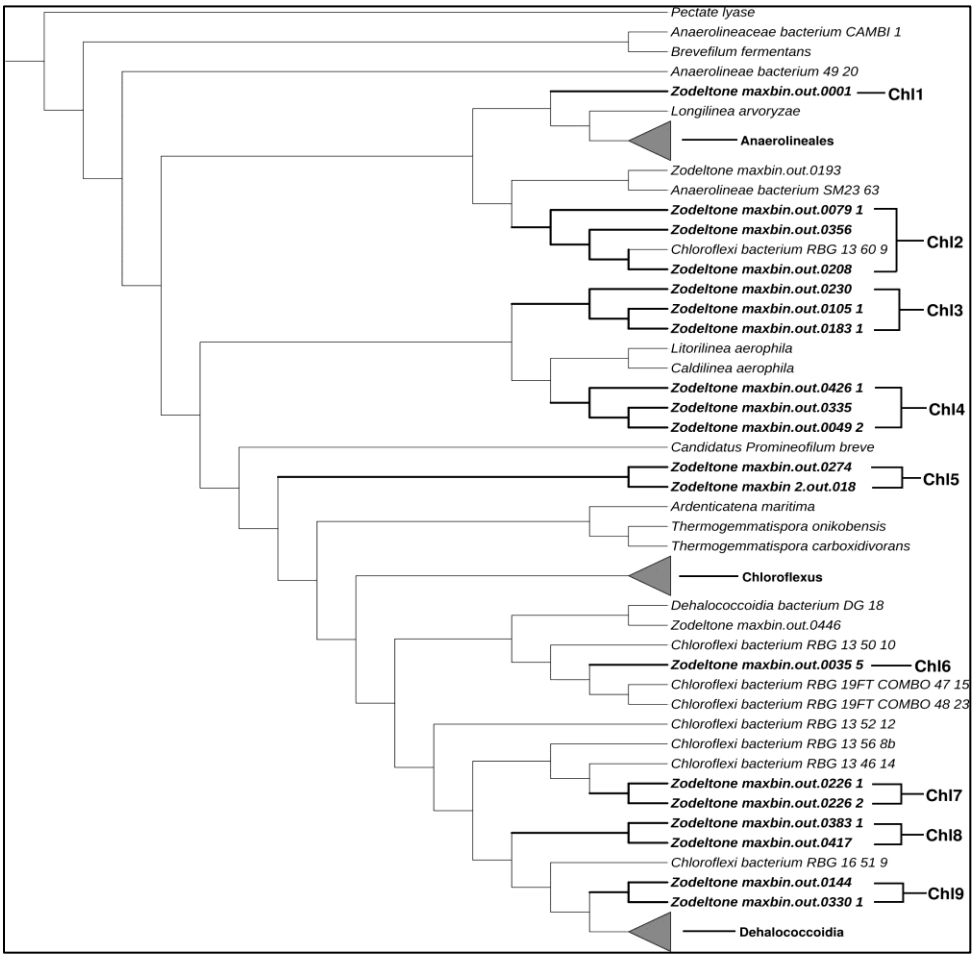

Figure 4. Phylogenetic tree of Chloroflexi-associated Zodletone RPL3 sequences and associated reference RPL3 sequences retrieved from the NCBI database.

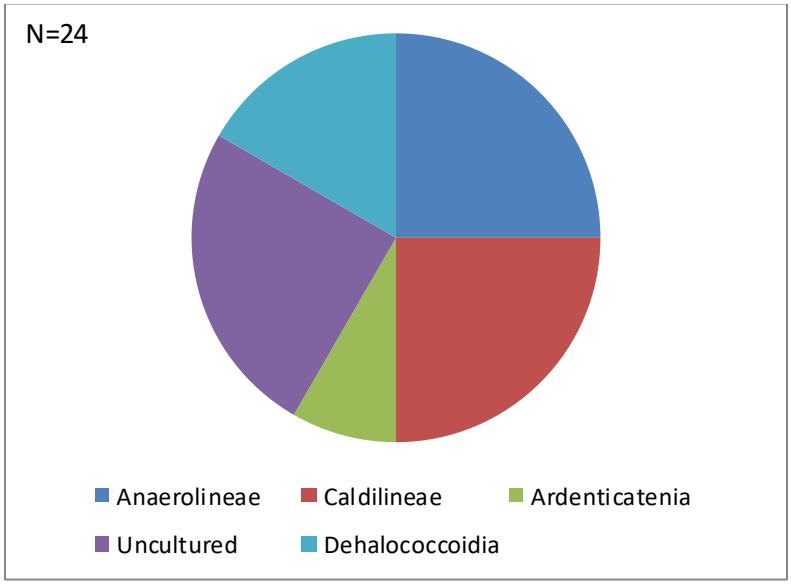

Figure 5. Phylogenetic assignments of Chloroflexi-associated Zodletone samples.

4. Clostridia.

6 of the 321 bacterial genomes were determined to belong to the class Clostridia through BLAST and phylogenetic tree affiliation (Table 3). Our focus was on the 3 genomes that formed 3 novel clades through phylogenetic affiliation, heretofore referred to as Clo1-3 (Figure 6). The other 3 genomes clustered with known sequences within Clostridia. The phylogenetic distribution can be seen in Figure 7.

Bin Name	Novel Clade Designation	Closest Relative	Closest Relative NCBI Accession	Identity Match	Putative Order
Zodletone_maxbin.out.0317_L3	Clo1	*Eubacterium uniforme*	WP_078765349.1	77%	Lachnospiraceae
Zodletone_maxbin.out.0284_1_L3	Clo2	*Caldicellulosiruptor morganii*	WP_045169206.1	58%	Thermoanaerobacterales
Zodletone_maxbin.out.0172_1_L3	Clo3	*Caloramator fervidus*	WP_103896573.1	55%	Uncultured
Zodletone_maxbin.out.0263_L3	N/A	*Tenericutes* bacterium HGW-*Tenericutes*-5	PKK94574.1	84%	Clostridiaceae
Zodletone_maxbin.out.0582_2_L3	N/A	*Firmicutes* bacterium HGW-*Firmicutes*-9	PKM40183.1	65%	Uncultured
Zodletone_maxbin.out.0469_L3	N/A	*Clostridium stercorarium*	WP_015360271.1	70%	Ruminococcaceae

Table 3. BLAST results of Clostridia-associated Zodletone samples and putative affiliation derived from clustering in phylogenetic tree.

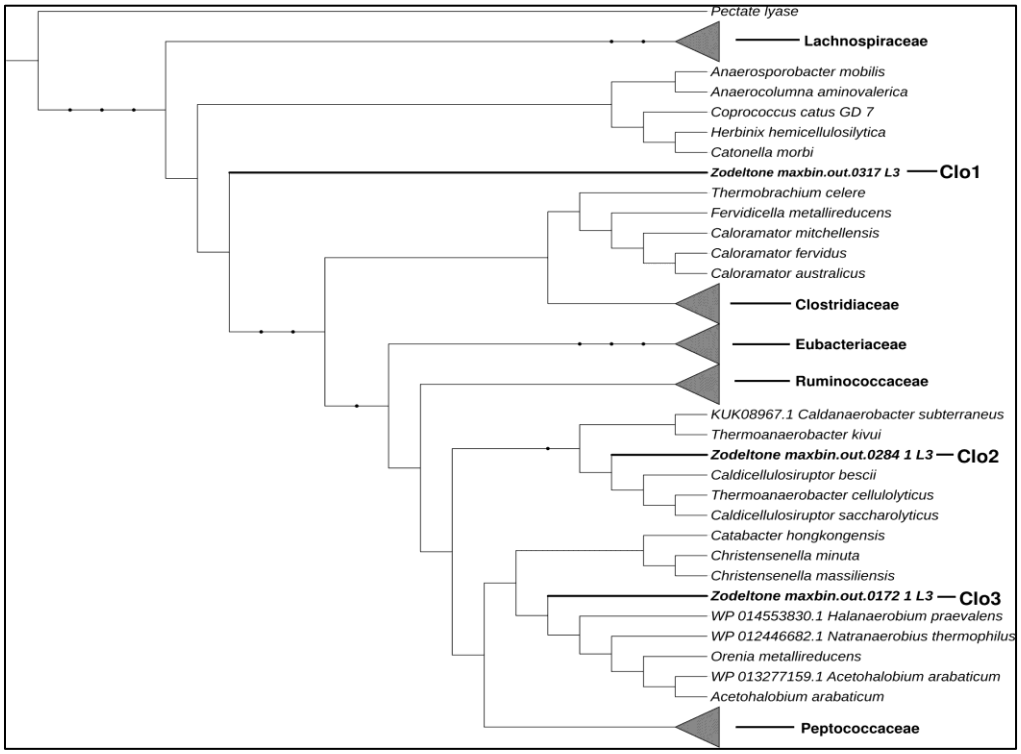

Figure 6. Phylogenetic tree of Clostridia-associated Zodletone RPL3 sequences and associated reference RPL3 sequences retrieved from the NCBI database.

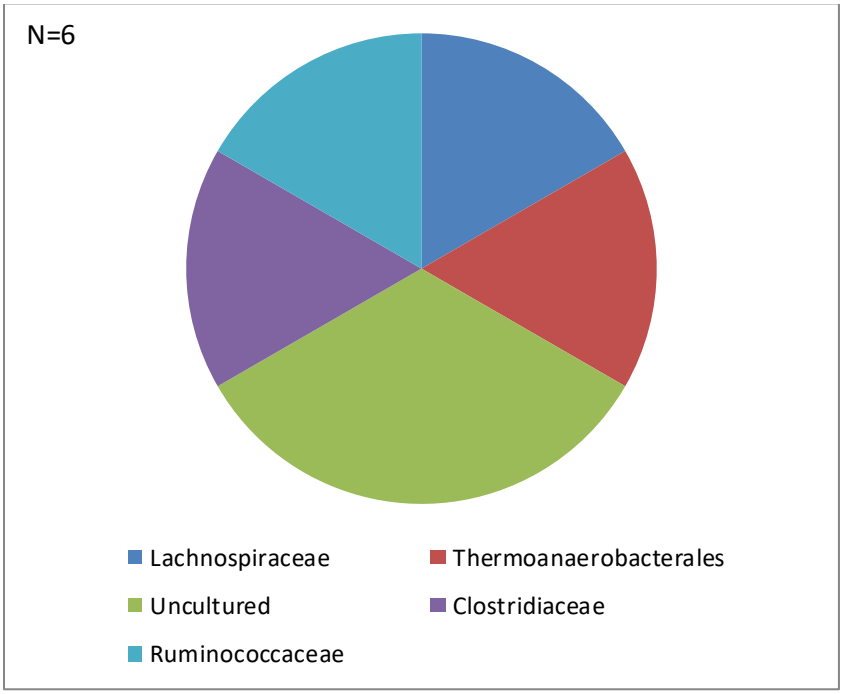

Figure 7. Phylogenetic assignments of Clostridia-associated Zodletone samples.

Discussion

Our results indicate that there are potentially novel organisms within Zodletone Spring, thus supporting our hypothesis. Previous research has indicated the presence of a diverse bacterial community within the Spring, notably sulfur- and sulfide reducing organisms. Previous studies have shown that Deltaproteobacteria and Chloroflexi composed a considerable portion of the microbial community, with Deltaproteobacteria composing 10% of the clone library of the microbial mat (Elshahed 2003). Springs similar to that of Zodletone have also had a notable population of Deltaproteobacteria and Chloroflexi (Chaudhary et al. 2009; Headd and Engel 2014). The lineages are related to diverse communities, including microbial communities isolated from the soil of Rifle, Colorado (Genbank PRJNA330071) and estuary sediments (Baker et al.).

Firmicutes have shown to be present in the Spring in previous studies, with their presence being detected with other nonphototrophic lineages (Elshahed 2003). Some Zodletone sequences clustered with known representatives of Clostridia. Others represented novel lineages. Clo1 appears to diverge at the order level, splitting the families Clostridiaceae and Lachnospiraceae. The similarity of its sequence is similar to that of both families, being 77% and 75% respectively. Clo2 split at the genus level, with matches with organisms of order Thermoanaerobacterales of ~58% and represents a potentially new genus within Family Thermoanaerobacterales. Clo3 diverged at the family level. It matched with organisms of several families, including Clostridiaceae and Halanaerobiales at 54% and 55% respectively, marking a potentially new family. In the future, we will further characterize the metabolic capabilities, physiological preferences, and ecological niches of these organisms.

Our research also shows that the RPL3 gene is capable of being use for the phylogenetic characterization of unknown organisms. The 16S approach is commonly used in bioinformatics, but carries several downsides namely the method has a tendency to create chimeric sequences and has low phylogenetic power at the species level and has difficulty discriminating between some genera (Mignard and Flandrois).

References

Baker, Brett J. et al. "Genomic Resolution of Linkages in Carbon, Nitrogen, and Sulfur Cycling among Widespread Estuary Sediment Bacteria." *Microbiome*, vol. 3, no. 1, 2015, p. 14, doi:10.1186/s40168-015-0077-6.

Bolger, Anthony M. et al. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics*, vol. 30, no. 15, 2014, pp. 2114-2120, doi:10.1093/bioinformatics/btu170.

Edgar, Robert C. "Muscle: Multiple Sequence Alignment with High Accuracy and High Throughput." *Nucleic Acids Research*, vol. 32, no. 5, 2004, pp. 1792-1797, PMC, doi:10.1093/nar/gkh340.

Elshahed, Mostafa S. et al. "Bacterial Diversity and Sulfur Cycling in a Mesophilic Sulfide-Rich Spring." *Applied and Environmental Microbiology*, vol. 69, no. 9, 2003, pp. 5609-5621, PMC, doi:10.1128/AEM.69.9.5609-5621.2003.

Finn, Robert D. et al. "Hmmer Web Server: Interactive Sequence Similarity Searching." *Nucleic Acids Research*, vol. 39, no. suppl_2, 2011, pp. W29-W37, doi:10.1093/nar/gkr367.

Hoehler, Tori M. and Bo Barker Jørgensen. "Microbial Life under Extreme Energy Limitation." *Nature Reviews Microbiology*, vol. 11, 2013, p. 83, doi:10.1038/nrmicro2939.

Joshi, N. A. and J. N. Fass. "Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for Fastq Files." 2011. doi:citeulike-article-id:13260426; http://github.com/najoshi/sickle.

Kang, Dongwan D. et al. "Metabat, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities." *PeerJ*, vol. 3, 2015, p. e1165, doi:10.7717/peerj.1165.

Laczny, Cedric C. et al. "Vizbin - an Application for Reference-Independent Visualization and Human-Augmented Binning of Metagenomic Data." *Microbiome*, vol. 3, no. 1, 2015, p. 1, doi:10.1186/s40168-014-0066-1.

Li, D. et al. "Megahit: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly Via Succinct De Bruijn Graph." *Bioinformatics*, vol. 31, no. 10, 2015, pp. 1674-1676, doi:10.1093/bioinformatics/btv033.

Luo, Qingwei et al. "Diversity of the Microeukaryotic Community in Sulfide-Rich Zodletone Spring (Oklahoma)." *Applied and Environmental Microbiology*, vol. 71, no. 10, 2005, pp. 6175-6184, PMC, doi:10.1128/AEM.71.10.6175-6184.2005.

Mardis, Elaine R. "Next-Generation Sequencing Platforms." *Annual Review of Analytical Chemistry*, vol. 6, no. 1, 2013, pp. 287-303, doi:10.1146/annurev-anchem-062012-092628.

Mignard, S. and J. P. Flandrois. "16s Rrna Sequencing in Routine Bacterial Identification: A 30-Month Experiment." *Journal of Microbiological Methods*, vol. 67, no. 3, 2006, pp. 574-581, doi:https://doi.org/10.1016/j.mimet.2006.05.009.

Parks, Donovan H. et al. "Checkm: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes." *Genome Research*, vol. 25, no. 7, 2015, pp. 1043-1055, PMC, doi:10.1101/gr.186072.114.

Rinke, Christian et al. "Insights into the Phylogeny and Coding Potential of Microbial Dark Matter." *Nature*, vol. 499, 2013, p. 431, doi:10.1038/nature12352 https://www.nature.com/articles/nature12352#supplementary-information.

Stamatakis, Alexandros. "Raxml Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics*, vol. 30, no. 9, 2014, pp. 1312-1313, doi:10.1093/bioinformatics/btu033.

Suzek, Baris E. et al. "Uniref Clusters: A Comprehensive and Scalable Alternative for Improving Sequence Similarity Searches." *Bioinformatics*, vol. 31, no. 6, 2015, pp. 926-932, PMC, doi:10.1093/bioinformatics/btu739.

Vollmers, John et al. "Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters!" *PLOS ONE*, vol. 12, no. 1, 2017, p. e0169662, doi:10.1371/journal.pone.0169662.

Waterhouse, Andrew M. et al. "Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench." *Bioinformatics*, vol. 25, no. 9, 2009, pp. 1189-1191, doi:10.1093/bioinformatics/btp033.

Wu, Yu-Wei et al. "Maxbin: An Automated Binning Method to Recover Individual Genomes from Metagenomes Using an Expectation-Maximization Algorithm." *Microbiome*, vol. 2, no. 1, 2014, p. 26, doi:10.1186/2049-2618-2-26.